Live Transcription using Web Speech API in JavaScript

Himanshi Jain Information Technology Department,

Thakur College of Engineering and Technology, Mumbai

Megha Jain Information Technology

Department,

Thakur College of Engineering and Technology, Mumbai

Sejal Jain Information Technology

Department,

Thakur College of Engineering and Technology, Mumbai

***Abstract: Holding meetings in the current pandemic has become an important part of students, employees and other staff. The conference covers a wide range of knowledge sharing and collaboration activities; including interviews, informal discussions, brainstorming sessions, video conferences and presentations. Therefore, we are interested in finding ways to improve the information obtained from these meetings. This report introduces the progress of using speech recognition technology to automatically generate text records of collaborative meetings. The report reviewed the literature on automatic transcription systems and described the research and development progress of the concept of automatic transcription. Other people with disabilities can also use this tool to help them communicate through meetings, video conferences, etc.***

1. INTRODUCTION

Live Transcription is a characteristic that inevitably transliterates speech in near real time, allowing people to connect in situations that they might not otherwise be able to. This feature uses voice recognition technology which enables voice input into the system. It is supposed to be the skill of a machine to identify words and phrases in the verbal language and then transform them into a machine-readable setup. In a nutshell, it means a computer program that learns to take input from human speech, which is then interpreted and ultimately written. Speech recognition is slowly becoming a part of our lives in the form of voice assistants like Alexa, Google Assistant, and Siri. Whether it's dictating words into your device to compose a document, searching the web by voice, or controlling your computer with your voice,

speech-to-text conversion makes their lives easier. We are faster and more comfortable. It has the potential to replace traditional forms of human-machine interface input devices, such as keyboards. A future where humans can interact with machines simply using voice and body movements is not far away. Speech recognition can be applied in the browser by means of JavaScript. Speech Recognition API allows a web application to accept voice input through the device's microphone and convert speech to text by matching words in the voice to words in its vocabulary.

Therefore, we used this technique to build a live transcriber that helps other students to record lectures in text format and create notes and other materials. Not only students but other people with disabilities i.e. hearing-impaired students can record lectures and other social conversations directed by the professor.

1. LITERATURE SURVEY

Luppi et al. (2009) argued that the introduction of Speech-to-Text Recognition (STR) technology in the traditional learning environment has several advantages, one of which is to improve teaching methods and improve learning opportunities. An approach of teaching students with different learning styles, which provides practical means for educators, to make their teaching accessible while improving the quality of teaching (Luppi et al., 2009).

Ranchal et al. (2013) introduced IBM ViaScribe and the IBM Hosted Transcription Service to assist college students with lecturing in the life and social sciences. The study evaluated two different methods of STR- mediated lecture recording, such as real-time subtitling and post-lecture transcription. The

instructor underwent initial voice training to develop a voice profile for the systems and improve STR accuracy before beginning real- time or post-conference transcription of closed captions. During the class, the STR processed verbal information into text signatures and transmitted it to a screen or to the students' computers. Any drafts received, unedited transcripts during the lectures, errors in the transcribed text were, however, corrected and made available to the students after the class.

Ryba et al. (2006) used the IBM ViaVoice and Viascribe systems to listen to business informatics lectures in a university lecture hall with more than 100 students. Initially, the professor was trained to develop a language profile for the system and achieve high precision. by introducing dialogue and vocabulary into the system. The STR systems then transliterated the lectures into texts that were showed on a big screen in front of the lecture hall. In addition, the edited texts generated by STR were sent to the students via the Internet after the conference. Ryba et al. (2006) examined the perception of students about the use of STR texts and the extent to which students use them. It also explored the main advantages and limitations of using STR texts.

Goddard, Kaplan, Kuehnle and Beglau (2007) implemented the Read and Write GOLD system as an intellectual prosthesis to provide the education needs of students in general tutoring classrooms. Teachers and students were initially trained in how to use the system. Students then used STR to achieve a certain understanding of the study material that they could not have understood before. Once students reached the level of understanding of their classmates, they reduced or stopped using the STR.

1. PROBLEM STATEMENT

With Covid19 Lockdown, currently all discussion and learning parts are done in online mode. Therefore, in most cases it will be difficult to make notes separately from the

presentation provided by the institution and other materials. Also, not only regular students but the impaired students with hearing disability face a lot of issues regarding communication via online meetings and video conferencing. Therefore, we need tools to help students get through their studies easily along with the students that are deaf and hard of hearing.

1. OBJECTIVES
2. During and after academic activities, help students understand the topics presented, catch up on missed or misheard parts of a speech, take notes, and do homework.
3. Allow students to pay more attention to the teacher instead of focusing on taking complete class notes and allow them to review the class material multiple times.
4. Help students with disabilities to better understand the content of the lectures and to improve their learning.
5. Helping hearing impaired students go through notes to fill the gaps in their understanding of what happened during the class.
6. MODULES AND FUNCTIONALITIES:

Here, the WebkitSpeechRecognition is used which is a Web Speech API object. It offers the speech interface, and established some of its attributes and event handlers. Below mentioned properties of WebKitSpeechRecognition object are used:

# Continuous listening

The speech recognition object can stop listening after the user stops talking, or it can continue listening until the user stops. In this project, this property is set to true so that even if the user pauses while speaking, the recognition will continue.

# Interim results

The intermediate result has not yet been finalized. If this property is permitted, the Speech Recognition object will also return transitional and ultimate results. Gray text is temporary and sometimes changed text, while black text is a recognizer response that is marked as final and will not change.

# Language

The language that the user will speak in. The use of locale codes to set this property was made.

# Events & callbacks

* + onStart

This event is triggered when the user starts voice recognition. A callback function is submitted to indicate that the voice recognition entity is monitoring the web page.

* + onEnd

This event is triggered when the user terminates speech recognition.

* + onError

This event is triggered if an error occurs during speech recognition.

* + onResult

This event is triggered when the SpeechRecognition object has some recognition results. It contains final and intermediate results.

Further, to activate the speech recognizer, recognition.start() function is called. Once it starts recording audio, it calls the onstart event handler, and then calls the onresult event handler for each new result set. This handler concatenates all the results received so far into two strings: final\_transcript and intermediate text.

recognition.start()

This method is used to start speech recognition. Once we start speech recognition, we can use the onstart event handler to notify users that speech recognition has started and that they should speak into the microphone. Usually this happens so quickly that the slash is imperceptible, but when using voice

recognition for the first time, Chrome must ask the user for permission to use the microphone. In this case, onstart will only be triggered when the user grants permission.

The project contains the following JavaScript functions:

* + createSession**:** This function creates a new session for each recording/transcription.
  + continueSession**:** If the user wants to continue the previous recording or session for any reason, this function provides the same possibility.
  + viewSession**:** This helps the user to select and view a specific session from the recorded sessions.
  + showRecordings**:** The recordings of the selected session will be displayed to the user.
  + recordSession**:** This function records the session in the background of real- time transcription, if the user continues the session, it will continue.
  + saveToLocalStorage**:** Even if the user exists the application, the sessions and recordings will be saved in the local storage. This avoids the loss of data or session recordings in the application.
  + download: As the feature implies, this feature enables users to edit recordings on the application and then save and download them to their device.

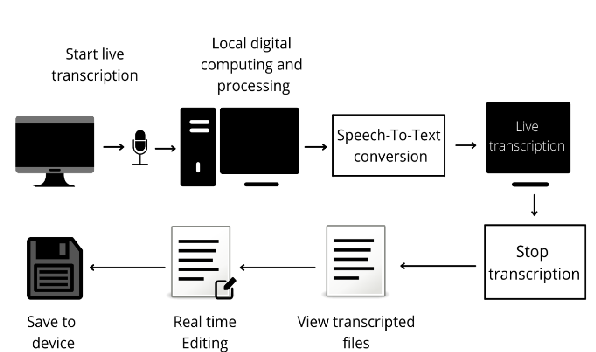
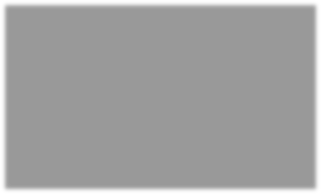
1. SOFTWARE AND HARDWARE

REQUIREMENTS:

JavaScript, HTML, CSS and WebkitSpeechRecognition

1. ARCHITECTURE DESIGN

technologies, such as automatic translation, to generate text from a speech and translate it into multiple languages at the same time. This method will provide teachers and students with instant audio-lingual interpretations in their native languages. It can be applied in a learning environment not only with a single speaker but with multi speakers as well.



1. RESULT AND DISCUSSION

Based on the advancement of STR technological development, STR technology was used to aid learning in a variety of ways. STR technology can help extend learning both during and after learning activities. There is widespread agreement on the numerous distinct advantages of STR-texts, such as allowing students to better understand the content of academic activities and to confirm missed parts of a speech, to take notes, to complete homework, and to prepare for exams. However, some debates about STR technology considerations that limit the technology's educational value persist. Students could benefit from both real-time and post-lecture transcriptions if they used this technology. Transcripts reduce memory workload by serving as external tokens for information that would otherwise be remembered. They act as visual retrieval cues for long-term memory, eliciting relevant information that would not otherwise be retrieved. Finally, transcripts are more “durable” (visual) text-based content that complements the more “temporary” (oral) speech-based presentation.

1. FUTURE SCOPE

This speech-to-text transcription technology can be expanded by combining it with other

REFERENCES

1. K. Bain, S. Basson, and M. Wald, “Speech Recognition in University Classrooms: Liberated Learning Project,” Proc. Fifth Int’l ACM SIGCAPH Conf. Assistive Technologies (ASSETS ’02), pp. 192-196, 2002.
2. K. Bain, S. Basson, A. Faisman, and D. Kanevsky, “Accessibility, Transcription, and Access Everywhere,” IBM Systems J., vol. 44, no. 3, pp. 589-604, 2005.
3. D. Leitch and T. MacMillan, “Liberated Learning Initiative Innovative Technology and Inclusion: Current Issues and Future Directions for Liberated Learning Research,” Year III Report, Saint Mary’s Univ., Nova Scotia, Canada, 2003.
4. M. Wald, “Synote: Accessible and Assistive Technology Enhancing Learning for All Students,” Proc. 12th Int’l Conf. Computers Helping People with Special Needs (ICCHP’10), pp. 177-184, 2010.
5. M. Wald, “Captioning for Deaf and Hard of Hearing People by Editing Automatic Speech Recognition in Real Time,” Proc. 10th Int’l Conf. Computers Helping People with Special Needs (ICCHP ’06), pp. 683-690, 2006.